



Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

A Generalization of the Borkar-Meyn Theorem for Stochastic Recursive Inclusions

Arunselvan Ramaswamy, Shalabh Bhatnagar

To cite this article:

Arunselvan Ramaswamy, Shalabh Bhatnagar (2017) A Generalization of the Borkar-Meyn Theorem for Stochastic Recursive Inclusions. *Mathematics of Operations Research* 42(3):648-661. <https://doi.org/10.1287/moor.2016.0821>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

A Generalization of the Borkar-Meyn Theorem for Stochastic Recursive Inclusions

Arunselvan Ramaswamy,^a Shalabh Bhatnagar^a

^aDepartment of Computer Science and Automation, Indian Institute of Science, Bangalore 560012, India

Contact: arunselvan@csa.iisc.ernet.in (AR); shalabh@csa.iisc.ernet.in (SB)

Received: February 26, 2015

Revised: December 31, 2015

Accepted: August 6, 2016

Published Online in Articles in Advance:
December 16, 2016

MSC2010 Subject Classification: Primary:
62L20; secondary: 93E03; 93E35; 34A60

OR/MS Subject Classification: Primary:
stochastic model applications; secondary:
analysis of algorithms

<https://doi.org/10.1287/moor.2016.0821>

Copyright: © 2016 INFORMS

Abstract. In this paper, the stability theorem of Borkar and Meyn is extended to include the case when the mean field is a set-valued map. Two different sets of sufficient conditions are presented that guarantee the “stability and convergence” of stochastic recursive inclusions. Our work builds on the works of Benaïm, Hofbauer and Sorin as well as Borkar and Meyn. As a corollary to one of the main theorems, a natural generalization of the Borkar and Meyn theorem follows. In addition, the original theorem of Borkar and Meyn is shown to hold under slightly relaxed assumptions. As an application to one of the main theorems, we discuss a solution to the “approximate drift problem.” Finally, we analyze the stochastic gradient algorithm with “constant-error gradient estimators” as yet another application of our main result.

Funding: This work was partly supported by the Robert Bosch Centre for Cyber-Physical Systems, Indian Institute of Science, Bangalore.

Keywords: stochastic recursive inclusions • Borkar-Meyn theorem • set-valued dynamical systems • ordinary differential equation method • stability criterion • stochastic gradient descent • iterative methods

1. Introduction

Consider the following recursion in \mathbb{R}^d ($d \geq 1$):

$$x_{n+1} = x_n + a(n)[h(x_n) + M_{n+1}], \quad \text{for } n \geq 0, \text{ where} \quad (1)$$

- (i) $h: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a Lipschitz-continuous function.
- (ii) $a(n) > 0$ for all n , is the stepsize sequence satisfying $\sum_{n=0}^{\infty} a(n) = \infty$ and $\sum_{n=0}^{\infty} a(n)^2 < \infty$.
- (iii) $M_n, n \geq 1$ is a sequence of martingale difference terms that constitute the noise.

The stochastic recursion given by (1) is often referred to as a stochastic recursive equation (SRE). A powerful method to analyze the limiting behavior of (1) is the ordinary differential equation (ODE) method. Here, the limiting behavior of the algorithm is described in terms of the asymptotics of the solution to the ODE

$$\dot{x}(t) = h(x(t)).$$

This method was introduced by Ljung [13] in 1977. For a detailed exposition on the subject and a survey of results, the reader is referred to Kushner and Yin [12] as well as Borkar [11].

In 1996, Benaïm [4] showed that the asymptotic behavior of a SRE can be studied by analyzing the asymptotic behavior of the associated ODE. However, no assumptions were made on the dynamics of the o.d.e. Specifically, Benaïm [4] developed sufficient conditions, which guarantee that limit sets of the continuously interpolated stochastic iterates are compact, connected, internally chain transitive and invariant sets of the associated o.d.e. The results found in Benaïm [4] are generalized in Benaïm [5]; further studies were made by Benaïm and Hirsch [6]. The assumptions made in Benaïm [4] are sometimes referred to as the “classical assumptions.” One of the key assumptions used by Benaïm to prove this convergence theorem is the almost sure boundedness of the iterates, i.e., stability of the iterates. In 1999, Borkar and Meyn [10] developed sufficient conditions, which guarantee the stability and convergence of stochastic recursive equations. These assumptions were consistent with those developed in Benaïm [4]. In this paper, we refer to the main result of Borkar and Meyn colloquially as the Borkar-Meyn theorem. In the same paper, Borkar and Meyn [10], several applications to problems from reinforcement learning have also been discussed. Another set of sufficient conditions for SREs were developed by Andrieu et al. [1] using global Lyapunov functions that guarantee the stability and convergence of the iterates.

In 2005, Benaïm et al. [7] showed that the dynamical systems approach can be extended to the situation where the mean fields are set valued. The algorithms considered were of the form:

$$x_{n+1} = x_n + a(n)[y_n + M_{n+1}] \quad \text{for } n \geq 0, \text{ where} \quad (2)$$

(i) $y_n \in h(x_n)$ and $h: \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ is a Marchaud map. For the definition of Marchaud maps, the reader is referred to Section 2.1.

(ii) $a(n) > 0$, for all $n \geq 0$, is the stepsize sequence satisfying $\sum_{n=0}^{\infty} a(n) = \infty$ and $\sum_{n=0}^{\infty} a(n)^2 < \infty$.

(iii) $M_n, n \geq 1$ is a sequence of martingale difference terms.

A recursion such as (2) is also called stochastic recursive inclusion (SRI). Since a differential equation can be seen as a special case of a differential inclusion wherein $h(x)$ is a cardinality one set for all $x \in \mathbb{R}^d$, SRE (1) can be seen as a special case of SRI (2).

The main aim of this paper is to extend the original Borkar-Meyn theorem to the case of stochastic recursive inclusions. We present two overlapping yet different sets of assumptions, in Sections 2.2 and 3.3, respectively, that guarantee the stability and convergence of an SRI given by (2). As a consequence of our main results, Theorems 2 and 3, we present a couple of interesting extensions to the original theorem of Borkar and Meyn in Section 4. Using the frameworks presented herein, we provide a solution to the problem of approximate drift in Section 5.1. For more details on the approximate drift problem, the reader is referred to Borkar [11]. In Section 6, we discuss the generality, ease of verifiability, and we also try to explain why the assumptions are “natural” in some sense.

Stochastic gradient descent (SGD) is an important method to find minima of (continuously) differentiable functions. When implementing the corresponding approximation algorithm (see (13) in Section 5.2) using gradient estimators, an error is made at each step in calculating the gradient of the objective function. Let's call this error the “approximation error.” This is the case when using gradient estimators such as Kiefer-Wolfowitz, simultaneous perturbation stochastic approximation (SPSA), and smoothed functional (SF) schemes, see Bhatnagar et al. [9]. Suppose the perturbation parameters of the aforementioned estimators are kept constant, then the “approximation error” is bounded by a constant that depends on the size of the perturbation parameters. We call such estimators constant-error gradient estimators. In Section 5.2, we analyze the stochastic gradient approximation algorithm that uses a constant-error gradient estimator. Using Theorem 3, we show that the iterates are stable and converge to a δ -neighborhood of the minimum set, for a specified $\delta(> 0)$. Essentially, our framework gives a threshold $\epsilon(\delta)$ for the “approximation error” so that the stochastic gradient approximation algorithm is stable and converges to a δ -neighborhood of the minimum set.

It is worth noting that prior to this paper, one could only claim that an SGD using constant-error gradient estimators will only converge to some neighborhood of the minimum set with high probability. In contrast, our framework guarantees almost sure convergence to a small neighborhood of the minimum set.

2. Preliminaries and Assumptions

2.1. Definitions and Notations

The definitions and notations used in this paper are similar to those in Benaïm et al. [7], Aubin and Cellina [2], Aubin and Frankowska [3], and Borkar [11]. In this section, we present a few for easy reference.

A set-valued map $h: \mathbb{R}^n \rightarrow \{\text{subsets of } \mathbb{R}^m\}$ is called a Marchaud map if it satisfies the following properties:

(i) For each $x \in \mathbb{R}^n$, $h(x)$ is convex and compact.

(ii) (Pointwise boundedness). For each $x \in \mathbb{R}^n$, $\sup_{w \in h(x)} \|w\| < K(1 + \|x\|)$ for some $K > 0$.

(iii) h is an upper semicontinuous map. We say that h is upper semicontinuous, if given sequences $\{x_n\}_{n \geq 1}$ (in \mathbb{R}^n) and $\{y_n\}_{n \geq 1}$ (in \mathbb{R}^m) with $x_n \rightarrow x$, $y_n \rightarrow y$ and $y_n \in h(x_n)$, $n \geq 1$, implies that $y \in h(x)$. In other words, the graph of h , $\{(x, y): y \in h(x), x \in \mathbb{R}^n\}$, is closed in $\mathbb{R}^n \times \mathbb{R}^m$.

Let H be a Marchaud map on \mathbb{R}^d . The differential inclusion (DI) given by

$$\dot{x} \in H(x) \quad (3)$$

is guaranteed to have at least one solution that is absolutely continuous. The reader is referred to Aubin and Cellina [2] for more details. We say that $x \in \Sigma$ if x is an absolutely continuous map that satisfies (3). The set-valued semiflow Φ associated with (3) is defined on $[0, +\infty) \times \mathbb{R}^d$ as $\Phi_t(x) = \{x(t) \mid x \in \Sigma, x(0) = x\}$. Let $B \times M \subset [0, +\infty) \times \mathbb{R}^k$ and define

$$\Phi_B(M) = \bigcup_{t \in B, x \in M} \Phi_t(x).$$

Let $M \subseteq \mathbb{R}^d$, the ω -limit set be defined by $\omega_\Phi(M) = \bigcap_{t \geq 0} \overline{\Phi_{[t, +\infty)}(M)}$. Similarly, the limit set of a solution x is given by $L(x) = \bigcap_{t \geq 0} \overline{x([t, +\infty))}$.

$M \subseteq \mathbb{R}^d$ is invariant if for every $x \in M$, there exists a trajectory, x , entirely in M with $x(0) = x$. In other words, $x \in \Sigma$ with $x(t) \in M$ for all $t \geq 0$.

Let $x \in \mathbb{R}^d$ and $A \subseteq \mathbb{R}^d$, then $d(x, A) := \inf\{\|a - y\| \mid y \in A\}$. We define the δ -open neighborhood of A by $N^\delta(A) := \{x \mid d(x, A) < \delta\}$. The δ -closed neighborhood of A is defined by $\bar{N}^\delta(A) := \{x \mid d(x, A) \leq \delta\}$. The open ball of radius r around the origin is represented by $B_r(0)$, while the closed ball is represented by $\bar{B}_r(0)$.

Internally Chain Transitive Set: $M \subseteq \mathbb{R}^d$ is said to be internally chain transitive if M is compact and for every $x, y \in M$, $\epsilon > 0$, and $T > 0$, we have the following: There exist Φ^1, \dots, Φ^n that are n solutions to the DI $\dot{x}(t) \in h(x(t))$, a sequence $x_1(=x), \dots, x_{n+1}(=y) \subset M$ and n real numbers t_1, t_2, \dots, t_n greater than T such that: $\Phi_{t_i}^i(x_i) \in N^\epsilon(x_{i+1})$ and $\Phi_{[0, t_i]}^i(x_i) \subset M$ for $1 \leq i \leq n$. The sequence $(x_1(=x), \dots, x_{n+1}(=y))$ is called an (ϵ, T) chain in M from x to y .

$A \subseteq \mathbb{R}^d$ is an attracting set if it is compact and there exists a neighborhood U such that for any $\epsilon > 0$, $\exists T(\epsilon) \geq 0$ with $\Phi_{[T(\epsilon), +\infty)}(U) \subset N^\epsilon(A)$. Such a U is called the fundamental neighborhood of A . In addition to being compact if the attracting set is also invariant, then it is called an attractor. The basin of attraction of A is given by $B(A) = \{x \mid \omega_\Phi(x) \subset A\}$. It is called Lyapunov stable if for all $\delta > 0$, $\exists \epsilon > 0$ such that $\Phi_{[0, +\infty)}(N^\epsilon(A)) \subseteq N^\delta(A)$. We use $T(\epsilon)$ and T_ϵ interchangeably to denote the dependence of T on ϵ .

We define the lower and upper limits of sequences of sets. Let $\{K_n\}_{n \geq 1}$ be a sequence of sets in \mathbb{R}^d .

1. The lower limit of $\{K_n\}_{n \geq 1}$ is given by $\text{Lim inf}_{n \rightarrow \infty} K_n := \{x \mid \lim_{n \rightarrow \infty} d(x, K_n) = 0\}$.

2. The upper limit of $\{K_n\}_{n \geq 1}$ is given by $\text{Lim sup}_{n \rightarrow \infty} K_n := \{y \mid \underline{\lim}_{n \rightarrow \infty} d(y, K_n) = 0\}$.

We may interpret that the lower limit collects the limit points of $\{K_n\}_{n \geq 1}$, while the upper limit collects its accumulation points.

2.2. The Assumptions

Recall that we have the following recursion in \mathbb{R}^d :

$$x_{n+1} = x_n + a(n)[y_n + M_{n+1}], \quad \text{where } y_n \in h(x_n).$$

We state our assumptions as follows:

(A1) $h: \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ is a Marchaud map.

(A2) $\{a(n)\}_{n \geq 0}$ is a scalar sequence such that $a(n) > 0 \forall n$, $\sum_{n \geq 0} a(n) = \infty$ and $\sum_{n \geq 0} a(n)^2 < \infty$. Without loss of generality, we let $\sup_n a(n) \leq 1$.

(A3) $\{M_n\}_{n \geq 1}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_n := \sigma(x_0, M_1, \dots, M_n)$, $n \geq 0$.

(i) $\{M_n\}_{n \geq 1}$ is a square integrable sequence.

(ii) $E[\|M_{n+1}\|^2 \mid \mathcal{F}_n] \leq K(1 + \|x_n\|^2)$ for $n \geq 0$ and some constant $K > 0$. Without loss of generality, assume that the same constant, K , works for both the pointwise boundedness condition of (A1) (see condition (ii) in the definition of Marchaud map in Section 2.1) and (A3).

For $c \geq 1$ and $x \in \mathbb{R}^d$, define $h_c(x) = \{y \mid cy \in h(cx)\}$. Further, for each $x \in \mathbb{R}^d$, define $h_\infty(x) := \overline{\text{Lim inf}_{c \rightarrow \infty} h_c(x)}$, i.e., the closure of the lower limit of $\{h_c(x)\}_{c \geq 1}$.

(A4) $h_\infty(x)$ is non empty for all $x \in \mathbb{R}^d$. Further, the differential inclusion $\dot{x}(t) \in h_\infty(x(t))$ has an attracting set, \mathcal{A} , with $\bar{B}_1(0)$ as a subset of its fundamental neighborhood. This attracting set is such that $\mathcal{A} \subseteq B_1(0)$.

(A5) Let $c_n \geq 1$ be an increasing sequence of integers such that $c_n \uparrow \infty$ as $n \rightarrow \infty$. Further, let $x_n \rightarrow x$ and $y_n \rightarrow y$ as $n \rightarrow \infty$ such that $y_n \in h_{c_n}(x_n)$, $\forall n$, then $y \in h_\infty(x)$.

Since the attracting set, $\mathcal{A} \subseteq B_1(0)$, is compact, we conclude that $\sup_{x \in \mathcal{A}} \|x\| < 1$. To see this, for all $x \in \mathcal{A}$, define $\delta(x) := \sup_{y \in \bar{B}_{\epsilon(x)}(x)} \|y\|$, where $\epsilon(x) > 0$ and $\bar{B}_{\epsilon(x)}(x) \subseteq B_1(0)$. For all $x \in \mathcal{A}$, we have $\delta(x) < 1$. Further, $\{B_{\epsilon(x)}(x) \mid x \in \mathcal{A}\}$ is an open cover of \mathcal{A} . Let $\{B_{\epsilon(x_i)}(x_i) \mid 1 \leq i \leq n\}$ be a finite subcover and $\delta := \max_{1 \leq i \leq n} \delta(x_i)$. Clearly, it follows that $\sup_{x \in \mathcal{A}} \|x\| \leq \delta < 1$. Define $\delta_1 := \sup_{x \in \mathcal{A}} \|x\|$ and pick real numbers δ_2, δ_3 , and δ_4 such that $\sup_{x \in \mathcal{A}} \|x\| = \delta_1 < \delta_2 < \delta_3 < \delta_4 < 1$. We shall use this sequence later on.

Assumptions (A1)–(A3) are the same as in Benaïm et al. [7]. However, the assumption on the stability of the iterates is replaced by (A4) and (A5). We show that (A4) and (A5) are sufficient conditions to ensure stability of iterates. We start by observing that h_c and h_∞ are Marchaud maps, where $c \geq 1$. Further, we show that the constant associated with the pointwise boundedness property is K of (A1) and (A3).

Proposition 1. h_∞ and h_c , $c \geq 1$, are Marchaud maps.

Proof. Fix $c \geq 1$ and $x \in \mathbb{R}^d$. To prove that $h_c(x)$ is compact, we show that it is closed and bounded. For $n \geq 1$, let $y_n \in h_c(x)$ and let $\lim_{n \rightarrow \infty} y_n = y$. It follows that $cy_n \in h(cx)$ for each $n \geq 1$ and $\lim_{n \rightarrow \infty} cy_n = cy$. Since $h(cx)$ is

closed, we have that $cy \in h(cx)$ and $y \in h_c(x)$. If we show that h_c is pointwise bounded, then we can conclude that $h_c(x)$ is compact. To prove the aforementioned, let $y \in h_c(x)$, then $cy \in h(cx)$. Since h satisfies (A1)(ii), we have that

$$c\|y\| \leq K(1 + \|cx\|), \quad \text{hence} \quad \|y\| \leq K\left(\frac{1}{c} + \|x\|\right).$$

Since $c (\geq 1)$ and x is arbitrarily chosen, h_c is pointwise bounded and the compactness of $h_c(x)$ follows. The set $h_c(x) = \{z/c \mid z \in h(cx)\}$ is convex since $h(cx)$ is convex and $h_c(x)$ is obtained by scaling it by $1/c$.

Next, we show that $h_c(x)$ is upper semicontinuous. Let $\lim_{n \rightarrow \infty} x_n = x$, $\lim_{n \rightarrow \infty} y_n = y$ and $y_n \in h_c(x_n)$, $\forall n \geq 1$. We need to show that $y \in h_c(x)$. We have that $cy_n \in h(cx_n)$ for each $n \geq 1$. Since $\lim_{n \rightarrow \infty} cx_n = cx$ and $\lim_{n \rightarrow \infty} cy_n = cy$, we conclude that $cy \in h(cx)$ since h is assumed to be upper semicontinuous.

It is left to show that $h_\infty(x)$, $x \in \mathbb{R}^d$ is a Marchaud map. To prove that $\|z\| \leq K(1 + \|x\|)$ for all $z \in h_\infty(x)$, it is enough to prove that $\|y\| \leq K(1 + \|x\|)$ for all $y \in \text{Lim inf}_{c \rightarrow \infty} h_c(x)$. Fix some $y \in \text{Lim inf}_{c \rightarrow \infty} h_c(x)$, then there exist $z_n \in h_n(x)$, $n \geq 1$ such that $\lim_{n \rightarrow \infty} \|y - z_n\| = 0$. We have that

$$\|y\| \leq \|y - z_n\| + \|z_n\|.$$

Since h_c , $c \geq 1$ is pointwise bounded (the constant associated is independent of c and equals K), the above inequality becomes

$$\|y\| \leq \|y - z_n\| + K(1 + \|x\|).$$

Letting $n \rightarrow \infty$ in the above inequality, we obtain $\|y\| \leq K(1 + \|x\|)$. Recall that $h_\infty(x) = \overline{\text{Lim inf}_{c \rightarrow \infty} h_c(x)}$, hence it is compact.

Again, to show that $h_\infty(x)$ is convex, for each $x \in \mathbb{R}^d$, we start by proving that $\text{Lim inf}_{c \rightarrow \infty} h_c(x)$ is convex. Let $u, v \in \text{Lim inf}_{c \rightarrow \infty} h_c(x)$ and $0 \leq t \leq 1$. We need to show that $tu + (1-t)v \in \text{Lim inf}_{c \rightarrow \infty} h_c(x)$. Consider an arbitrary sequence $\{c_n\}_{n \geq 1}$ such that $c_n \rightarrow \infty$, then there exist $u_n, v_n \in h_{c_n}(x)$ such that $\|u_n - u\|$ and $\|v_n - v\| \rightarrow 0$ as $c_n \rightarrow \infty$. Since $h_{c_n}(x)$ is convex, it follows that $tu_n + (1-t)v_n \in h_{c_n}(x)$, further,

$$\lim_{c_n \rightarrow \infty} (tu_n + (1-t)v_n) = tu + (1-t)v.$$

Since we started with an arbitrary sequence $c_n \rightarrow \infty$, it follows that $tu + (1-t)v \in \text{Lim inf}_{c \rightarrow \infty} h_c(x)$. Now, we can prove that $h_\infty(x)$ is convex. Let $u, v \in h_\infty(x)$. Then, $\exists \{u_n\}_{n \geq 1}$ and $\{v_n\}_{n \geq 1} \subseteq \text{Lim inf}_{c \rightarrow \infty} h_c(x)$ such that $u_n \rightarrow u$ and $v_n \rightarrow v$ as $n \rightarrow \infty$. We need to show that $tu + (1-t)v \in h_\infty(x)$ for $0 \leq t \leq 1$. Since $tu_n + (1-t)v_n \in \text{Lim inf}_{c \rightarrow \infty} h_c(x)$, the desired result is obtained by letting $n \rightarrow \infty$ in $tu_n + (1-t)v_n$.

Finally, we show that h_∞ is upper semicontinuous. Let $\lim_{n \rightarrow \infty} x_n = x$, $\lim_{n \rightarrow \infty} y_n = y$, and $y_n \in h_\infty(x_n)$, $\forall n \geq 1$. We need to show that $y \in h_\infty(x)$. Since $y_n \in h_\infty(x_n)$, $\exists z_n \in \text{Lim inf}_{c \rightarrow \infty} h_c(x_n)$ such that $\|z_n - y_n\| < 1/n$. Since $z_n \in \text{Lim inf}_{c \rightarrow \infty} h_c(x_n)$, $n \geq 1$, it follows that there exist c_n such that for all $c \geq c_n$, $d(z_n, h_c(x_n)) < 1/n$. In particular, $\exists u_n \in h_{c_n}(x_n)$ such that $\|z_n - u_n\| < 1/n$. We choose the sequence $\{c_n\}_{n \geq 1}$ such that $c_{n+1} > c_n$ for each $n \geq 1$. We now have the following: $\lim_{n \rightarrow \infty} u_n = y$, $u_n \in h_{c_n}(x_n) \forall n$ and $\lim_{n \rightarrow \infty} x_n = x$. It follows directly from assumption (A5) that $y \in h_\infty(x)$. \square

3. Stability and Convergence of Stochastic Recursive Inclusions

We begin by providing a brief outline of our approach to prove the stability of an SRI under assumptions (A1)–(A5). First, we divide the time line, $[0, \infty)$, approximately into intervals of length T . We shall explain later how we choose and fix T . Then, we construct a linearly interpolated trajectory from the given SRI; the construction is explained in the next paragraph. A sequence of “rescaled” trajectories of length T is constructed as follows: At the beginning of each T -length interval, we observe the trajectory to see if it is outside the unit ball, if so, we scale it back to the boundary of the unit ball. This scaling factor is then used to scale the “rest of the T -length trajectory.”

To show that the iterates are bounded almost surely, we need to show that the linearly interpolated trajectory does not “run off” to infinity. To do so, we assume that this is not true and show that there exists a subsequence of the rescaled T -length trajectories that has a solution to $\dot{x}(t) \in h_\infty(x(t))$ as a limit point in $C([0, T], \mathbb{R}^d)$. We choose and fix T such that any solution to $\dot{x}(t) \in h_\infty(x(t))$ with an initial value inside the unit ball is close to the origin at the end of time T . In this paper, we choose $T = T(\delta_2 - \delta_1) + 1$. We then argue that the linearly interpolated trajectory is forced to make arbitrarily large “jumps” within time T . The Gronwall inequality is then used to show that this is not possible.

Once we prove stability of the recursion, we invoke Theorem 3.6 and Lemma 3.8 from Benaïm et al. [7] to conclude that the limit set is a closed, connected, internally chain transitive and invariant set associated with $\dot{x}(t) \in h_\infty(x(t))$.

We construct the linearly interpolated trajectory $\bar{x}(t)$ for $t \in [0, \infty)$ from the sequence $\{x_n\}$ as follows: Define $t(0) := 0$, $t(n) := \sum_{i=0}^{n-1} a(i)$. Let $\bar{x}(t(n)) := x_n$ and for $t \in (t(n), t(n+1))$, let

$$\bar{x}(t) := \left(\frac{t(n+1) - t}{t(n+1) - t(n)} \right) \bar{x}(t(n)) + \left(\frac{t - t(n)}{t(n+1) - t(n)} \right) \bar{x}(t(n+1)).$$

We define a piecewise constant trajectory using the sequence $\{y_n\}_{n \geq 0}$ as follows: $\bar{y}(t) := y_n$ for $t \in [t(n), t(n+1))$, $n \geq 0$.

We know that the DI given by $\dot{x}(t) \in h_\infty(x(t))$ has an attractor set \mathcal{A} such that $\delta_1 := \sup_{x \in \mathcal{A}} \|x\| < 1$. Let us fix $T := T(\delta_2 - \delta_1) + 1$, where $T(\delta_2 - \delta_1)$ is as defined in Section 2.1. Then, $\|x(t)\| < \delta_2$ for all $t \geq T(\delta_2 - \delta_1)$, where $\{x(t) : t \in [0, \infty)\}$ is a solution to $\dot{x}(t) \in h_\infty(x(t))$ with an initial value inside the unit ball around the origin.

Define $T_0 := 0$ and $T_n := \min\{t(m) : t(m) \geq T_{n-1} + T\}$, $n \geq 1$. Observe that there exists a subsequence $\{m(n)\}_{n \geq 0}$ of $\{n\}$ such that $T_n = t(m(n)) \forall n \geq 0$. We construct the rescaled trajectory, $\hat{x}(t)$, $t \geq 0$, as follows: Let $t \in [T_n, T_{n+1})$ for some $n \geq 0$, then $\hat{x}(t) := \bar{x}(t)/(r(n))$, where $r(n) = \|\bar{x}(T_n)\| \vee 1$. Also, let $\hat{x}(T_{n+1}^-) := \lim_{t \uparrow T_{n+1}} \hat{x}(t)$, $t \in [T_n, T_{n+1})$. The corresponding “rescaled y iterates” are given by $\hat{y}(t) := \bar{y}(t)/(r(n))$ and the rescaled martingale noise terms by $\hat{M}_{k+1} := M_{k+1}/(r(n))$, $t(k) \in [T_n, T_{n+1})$, $n \geq 0$.

Consider the recursion at hand, i.e.,

$$\bar{x}(t(k+1)) = \bar{x}(t(k)) + a(k)(\bar{y}(t(k)) + M_{k+1})$$

such that $t(k), t(k+1) \in [T_n, T_{n+1})$. Multiplying both sides by $1/r(n)$, we get the rescaled recursion:

$$\hat{x}(t(k+1)) = \hat{x}(t(k)) + a(k)(\hat{y}(t(k)) + \hat{M}_{k+1}).$$

Since $\bar{y}(t(k)) \in h(\bar{x}(t(k)))$, it follows that $\hat{y}(t(k)) \in h_{r(n)}(\hat{x}(t(k)))$. It is worth noting that $E[\|\hat{M}_{k+1}\|^2 | \mathcal{F}_k] \leq K(1 + \|\hat{x}(t(k))\|^2)$.

3.1. Characterizing Limits, in $C([0, T], \mathbb{R}^d)$, of the Rescaled Trajectories

We do not provide proofs for the first three lemmas since they can be found in Borkar [11] or Benaïm et al. [7]. The first two lemmas essentially state that the rescaled martingale noise converges almost surely.

Lemma 1. $\sup_{t \in [0, T]} E\|\hat{x}(t)\|^2 < \infty$.

Lemma 2. The rescaled sequence $\{\hat{c}_n\}_{n \geq 1}$, where $\hat{c}_n = \sum_{k=0}^{n-1} a(k)\hat{M}_{k+1}$ is convergent almost surely.

The rest of the lemmas are needed to prove the stability theorem, Theorem 1. We begin by showing that the rescaled trajectories are bounded almost surely.

Lemma 3. $\sup_{t \in [0, \infty)} \|\hat{x}(t)\| < \infty$ a.s.

As stated earlier, we omit the proof of the above stated lemma and establish a couple of notations used later. Let $A = \{\omega \mid \{\hat{c}_n(\omega)\}_{n \geq 1} \text{ converges}\}$. Since \hat{c}_n , $n \geq 1$, converges on A , there exists $M_\omega < \infty$, possibly sample-path dependent, such that $\|\sum_{l=0}^{k-1} a(m(n) + l)\hat{M}_{m(n)+l+1}\| \leq M_\omega$, where M_ω is independent of n and k . Also, let $\sup_{t \geq 0} \|\hat{x}(t)\| \leq K_\omega$, where $K_\omega := (1 + M_\omega + (T + 1)K)e^{K(T+1)}$ is also a constant that is sample-path dependent.

Let $x^n(t)$, $t \in [0, T]$ be the solution (up to time T) to $\dot{x}^n(t) = \hat{y}(T_n + t)$, with the initial condition $x^n(0) = \hat{x}(T_n)$. Clearly, we have

$$x^n(t) = \hat{x}(T_n) + \int_0^t \hat{y}(T_n + z) dz. \tag{4}$$

The following two lemmas are inspired by ideas from Benaïm et al. [7] as well as Borkar [11]. In the lemma that follows, we show that the limit sets of $\{x^n(\cdot) \mid n \geq 0\}$ and $\{\hat{x}(T_n + \cdot) \mid n \geq 0\}$ coincide. We seek limits in $C([0, T], \mathbb{R}^d)$.

Lemma 4. $\lim_{n \rightarrow \infty} \sup_{t \in [T_n, T_n + T]} \|x^n(t) - \hat{x}(t)\| = 0$ a.s.

Proof. Let $t \in [t(m(n) + k), t(m(n) + k + 1))$ and $t(m(n) + k + 1) \leq T_{n+1}$. We first assume that $t(m(n) + k + 1) < T_{n+1}$. We have the following:

$$\hat{x}(t) = \left(\frac{t(m(n) + k + 1) - t}{a(m(n) + k)} \right) \hat{x}(t(m(n) + k)) + \left(\frac{t - t(m(n) + k)}{a(m(n) + k)} \right) \hat{x}(t(m(n) + k + 1)).$$

Substituting for $\hat{x}(t(m(n) + k + 1))$ in the above equation, we get

$$\begin{aligned} \hat{x}(t) &= \left(\frac{t(m(n) + k + 1) - t}{a(m(n) + k)} \right) \hat{x}(t(m(n) + k)) \\ &\quad + \left(\frac{t - t(m(n) + k)}{a(m(n) + k)} \right) (\hat{x}(t(m(n) + k)) + a(m(n) + k)(\hat{y}(t(m(n) + k)) + \hat{M}_{m(n)+k+1})), \end{aligned}$$

hence

$$\hat{x}(t) = \hat{x}(t(m(n) + k)) + (t - t(m(n) + k))(\hat{y}(t(m(n) + k)) + \hat{M}_{m(n)+k+1}).$$

Unfolding $\hat{x}(t(m(n) + k))$ over k , we get

$$\begin{aligned} \hat{x}(t) &= \hat{x}(T_n) + \sum_{l=0}^{k-1} a(m(n) + l)(\hat{y}(t(m(n) + l)) + \hat{M}_{m(n)+l+1}) \\ &\quad + (t - t(m(n) + k))(\hat{y}(t(m(n) + k)) + \hat{M}_{m(n)+k+1}). \end{aligned} \tag{5}$$

Now, we consider $x^n(t)$, i.e.,

$$x^n(t) = \hat{x}(T_n) + \int_0^t \hat{y}(T_n + z) dz.$$

Splitting the above integral, we get

$$x^n(t) = \hat{x}(T_n) + \sum_{l=0}^{k-1} \int_{t(m(n)+l}^{t(m(n)+l+1)} \hat{y}(z) dz + \int_{t(m(n)+k}^t \hat{y}(z) dz.$$

Thus

$$x^n(t) = \hat{x}(T_n) + \sum_{l=0}^{k-1} a(m(n) + l)\hat{y}(t(m(n) + l)) + (t - t(m(n) + k))\hat{y}(t(m(n) + k)). \tag{6}$$

From (5) and (6), it follows that:

$$\|x^n(t) - \hat{x}(t)\| \leq \left\| \sum_{l=0}^{k-1} a(m(n) + l)\hat{M}_{m(n)+l+1} \right\| + \|(t - t(m(n) + k))\hat{M}_{m(n)+k+1}\|,$$

and hence

$$\|x^n(t) - \hat{x}(t)\| \leq \|\hat{\mathcal{C}}_{m(n)+k} - \hat{\mathcal{C}}_{m(n)}\| + \|\hat{\mathcal{C}}_{m(n)+k+1} - \hat{\mathcal{C}}_{m(n)+k}\|.$$

If $t(m(n) + k + 1) = T_{n+1}$, then in the proof, we may replace $\hat{x}(t(m(n) + k + 1))$ with $\hat{x}(T_{n+1}^-)$. The arguments remain the same. Since $\hat{\mathcal{C}}_n, n \geq 1$ converges almost surely, the desired result follows. \square

The sets $\{x^n(t), t \in [0, T] \mid n \geq 0\}$ and $\{\hat{x}(T_n + t), t \in [0, T] \mid n \geq 0\}$ can be viewed as subsets of $C([0, T], \mathbb{R}^d)$. It can be shown that $\{x^n(t), t \in [0, T] \mid n \geq 0\}$ is equicontinuous and pointwise bounded. Thus from the Arzela-Ascoli theorem, $\{x^n(t), t \in [0, T] \mid n \geq 0\}$ is relatively compact. It follows from Lemma 4 that the set $\{\hat{x}(T_n + t), t \in [0, T] \mid n \geq 0\}$ is also relatively compact in $C([0, T], \mathbb{R}^d)$.

Lemma 5. Let $r(n) \uparrow \infty$, then any limit point of $\{\hat{x}(T_n + t), t \in [0, T]; n \geq 0\}$ is of the form $x(t) = x(0) + \int_0^t y(s) ds$, where $y: [0, T] \rightarrow \mathbb{R}^d$ is a measurable function and $y(t) \in h_\infty(x(t)), t \in [0, T]$.

Proof. We define the notation $[t] := \max\{t(k) \mid t(k) \leq t\}, t \geq 0$. Let $t \in [T_n, T_{n+1})$, then $\hat{y}(t) \in h_{r(n)}(\hat{x}([t]))$ and $\|\hat{y}(t)\| \leq K(1 + \|\hat{x}([t])\|)$ since $h_{r(n)}$ is a Marchaud map (K is the constant associated with the pointwise boundedness property). It follows from Lemma 3 that $\sup_{t \in [0, \infty)} \|\hat{y}(t)\| < \infty$ a.s. Using observations made earlier, we can deduce that there exists a subsequence of \mathbb{N} ; say, $\{l\} \subseteq \{n\}$ such that $\hat{x}(T_l + t) \rightarrow x(t)$ in $C([0, T], \mathbb{R}^d)$ and $\hat{y}(m(l) + \cdot) \rightarrow y(\cdot)$ weakly in $L_2([0, T], \mathbb{R}^d)$. From Lemma 4, it follows that $x^l(\cdot) \rightarrow x(\cdot)$ in $C([0, T], \mathbb{R}^d)$. Letting $r(l) \uparrow \infty$ in

$$x^l(t) = x^l(0) + \int_0^t \hat{y}(t(m(l) + z)) dz, \quad t \in [0, T],$$

we get $x(t) = x(0) + \int_0^t y(z) dz$ for $t \in [0, T]$. Since $\|\hat{x}(T_n)\| \leq 1$, we have $\|x(0)\| \leq 1$.

Since $\hat{y}(T_l + \cdot) \rightarrow y(\cdot)$ weakly in $L_2([0, T], \mathbb{R}^d)$, there exists $\{l(k)\} \subseteq \{l\}$ such that

$$\frac{1}{N} \sum_{k=1}^N \hat{y}(T_{l(k)} + \cdot) \rightarrow y(\cdot) \quad \text{strongly in } L_2([0, T], \mathbb{R}^d).$$

Further, there exists $\{N(m)\} \subseteq \{N\}$ such that

$$\frac{1}{N(m)} \sum_{k=1}^{N(m)} \hat{y}(T_{l(k)} + \cdot) \rightarrow y(\cdot) \quad \text{a.e. on } [0, T].$$

Let us fix $t_0 \in \{t \mid (1/N(m)) \sum_{k=1}^{N(m)} \hat{y}(T_{l(k)} + t) \rightarrow y(t), t \in [0, T]\}$, then

$$\lim_{N(m) \rightarrow \infty} \frac{1}{N(m)} \sum_{k=1}^{N(m)} \hat{y}(T_{l(k)} + t_0) = y(t_0).$$

Since $h_\infty(x(t_0))$ is convex and compact (Proposition 1), to show that $y(t_0) \in h_\infty(x(t_0))$, it is enough to prove that $\lim_{l(k) \rightarrow \infty} d(\hat{y}(T_{l(k)} + t_0), h_\infty(x(t_0))) = 0$. If not, $\exists \epsilon > 0$ and $\{n(k)\} \subseteq \{l(k)\}$ such that $d(\hat{y}(T_{n(k)} + t_0), h_\infty(x(t_0))) > \epsilon$. Since $\{\hat{y}(T_{n(k)} + t_0)\}_{k \geq 1}$ is norm bounded, it follows that there is a convergent subsequence. For the sake of convenience, we assume that $\lim_{k \rightarrow \infty} \hat{y}(T_{n(k)} + t_0) = y$ for some $y \in \mathbb{R}^d$. Since $\hat{y}(T_{n(k)} + t_0) \in h_{r(n(k))}(\hat{x}([T_{n(k)} + t_0]))$ and $\lim_{k \rightarrow \infty} \hat{x}([T_{n(k)} + t_0]) = x(t_0)$, it follows from assumption (A5) that $y \in h_\infty(x(t_0))$. This leads to a contradiction. \square

Note that in the statement of Lemma 5 we can replace “ $r(n) \uparrow \infty$ ” by “ $r(l) \uparrow \infty$,” where $\{r(l)\}$ is a subsequence of $\{r(n)\}$. Specifically, we can conclude that any limit point of $\{\hat{x}(T_k + t), t \in [0, T]\}_{\{k\} \subseteq \{n\}}$ in $C([0, T], \mathbb{R}^d)$, conditioned on $r(k) \uparrow \infty$, is of the form $x(t) = x(0) + \int_0^t y(z) dz$, where $y(t) \in h_\infty(x(t))$ for $t \in [0, T]$. It should be noted that $y(\cdot)$ may be sample-path dependent. The following is an immediate consequence of Lemma 5.

Corollary 1. $\exists R_0, 1 < R_0 < \infty$ such that $\forall r(l) > R_0, \|\hat{x}(T_l + \cdot) - x(\cdot)\| < \delta_3 - \delta_2$, where $\{l\} \subseteq \mathbb{N}$ and $x(\cdot)$ is a solution (up to time T) of $\dot{x}(t) \in h_\infty(x(t))$ such that $\|x(0)\| \leq 1$. The form of $x(\cdot)$ is as given by Lemma 5.

Proof. Assume to the contrary that $\exists r(l) \uparrow \infty$ such that $\hat{x}(T_l + \cdot)$ is at least $\delta_3 - \delta_2$ away from any solution to the DI. It follows from Lemma 5 that there exists a subsequence of $\{\hat{x}(T_l + t), 0 \leq t \leq T: l \subseteq \mathbb{N}\}$ guaranteed to converge, in $C([0, T], \mathbb{R}^d)$, to a solution of $\dot{x}(t) \in h_\infty(x(t))$ such that $\|x(0)\| \leq 1$. This is a contradiction. \square

It is worth noting that R_0 may be sample-path dependent. Since $T = T(\delta_2 - \delta_1) + 1$, we get $\|\hat{x}([T_l + T])\| < \delta_3$ for all T_l such that $\|\hat{x}(T_l)\| (= r(l)) > R_0$.

3.2. Stability Theorem

We are now ready to prove the stability of an SRI given by (2) under the assumptions (A1)–(A5). If $\sup_n r(n) < \infty$, then the iterates are stable and there is nothing to prove. If, on the other hand, $\sup_n r(n) = \infty$, there exists $\{l\} \subseteq \{n\}$ such that $r(l) \uparrow \infty$. It follows from Lemma 5 that any limit point of $\{\hat{x}(T_l + t), t \in [0, T]: \{l\} \subseteq \{n\}\}$ is of the form $x(t) = x(0) + \int_0^t y(s) ds$, where $y(t) \in h_\infty(x(t))$ for $t \in [0, T]$. From assumption (A4), we have that $\|x(T)\| < \delta_2$. Since the time intervals are roughly T apart, for large values of $r(n)$, we can conclude that $\|\hat{x}(T_{n+1}^-)\| < \delta_3$, where $\hat{x}(T_{n+1}^-) = \lim_{t \uparrow t(m(n+1))} \hat{x}(t), t \in [T_n, T_{n+1})$.

Theorem 1 (Stability Theorem for DI). *Under assumptions (A1)–(A5), $\sup_n \|x_n\| < \infty$ a.s.*

Proof. As explained earlier, it is sufficient to consider the case when $\sup_n r(n) = \infty$. Let $\{l\} \subseteq \{n\}$ such that $r(l) \uparrow \infty$. Recall that $T_l = t(m(l))$ and that $[T_l + T] = \max\{t(k) \mid t(k) \leq T_l + T\}$.

We have $\|x(T)\| < \delta_2$ since $x(t)$ is a solution, up to time T , to the DI given by $\dot{x}(t) \in h_\infty(x(t))$ and we have fixed $T = T(\delta_2 - \delta_1) + 1$. From Lemma 5, we conclude that there exists N such that all of the following happen:

- (i) $m(l) \geq N \Rightarrow \|\hat{x}([T_l + T])\| < \delta_3$.
- (ii) $n \geq N \Rightarrow a(n) < (\delta_4 - \delta_3)/[K(1 + K_\omega) + M_\omega]$.
- (iii) $n > m \geq N \Rightarrow \|\hat{c}_n - \hat{c}_m\| < M_\omega$.
- (iv) $m(l) \geq N \Rightarrow r(l) > R_0$.

In the above, R_0 is defined in the statement of Corollary 1 and K_ω, M_ω are explained in Lemma 3.

Recall that we chose $\sup_{x \in S_l} \|x\| = \delta_1 < \delta_2 < \delta_3 < \delta_4 < 1$ in Section 2.2. Let $m(l) \geq N$ and $t(m(l+1)) = t(m(l) + k + 1)$ for some $k \geq 0$. Clearly, from the manner in which the T_n sequence is defined, we have $t(m(l) + k) = [T_l + T]$. As defined earlier, $\hat{x}(T_{n+1}^-) = \lim_{t \uparrow t(m(n+1))} \hat{x}(t), t \in [T_n, T_{n+1})$ and $n \geq 0$. Consider the equation

$$\hat{x}(T_{l+1}^-) = \hat{x}(t(m(l) + k)) + a(m(l) + k)(\hat{y}(t(m(l) + k)) + \hat{M}_{m(l)+k+1}).$$

Taking norms on both sides, we get

$$\|\hat{x}(T_{l+1}^-)\| \leq \|\hat{x}(t(m(l) + k))\| + a(m(l) + k)\|\hat{y}(t(m(l) + k))\| + a(m(l) + k)\|\hat{M}_{m(l)+k+1}\|.$$

From the way we have chosen N , we conclude that

$$\|\hat{y}(t(m(l) + k))\| \leq K(1 + \|\hat{x}(t(m(l) + k))\|) \leq K(1 + K_\omega) \quad \text{and that}$$

$$\|\hat{M}_{m(l)+k+1}\| = \|\hat{C}_{m(l)+k+1} - \hat{C}_{m(l)+k}\| \leq M_\omega.$$

Thus we have that

$$\|\hat{x}(T_{l+1}^-)\| \leq \|\hat{x}(t(m(l) + k))\| + a(m(l) + k)(K(1 + K_\omega) + M_\omega).$$

Finally, we have that $\|\hat{x}(T_{l+1}^-)\| < \delta_4$ and

$$\frac{\|\bar{x}(T_{l+1})\|}{\|\bar{x}(T_l)\|} = \frac{\|\hat{x}(T_{l+1}^-)\|}{\|\hat{x}(T_l)\|} < \delta_4 < 1. \tag{7}$$

It follows from (7) that $\|\bar{x}(T_{n+1})\| < \delta_4\|\bar{x}(T_n)\|$ if $\|\bar{x}(T_n)\| > R_0$. From Corollary 1 and the aforementioned we get that the trajectory falls at an exponential rate till it enters $\bar{B}_{R_0}(0)$. Let $t \leq T_l$, $t \in [T_n, T_{n+1})$ and $n + 1 \leq l$ be the last time that $\bar{x}(t)$ jumps from $\bar{B}_{R_0}(0)$ to the outside of the ball. It follows that $\|\bar{x}(T_{n+1})\| \geq \|\bar{x}(T_l)\|$. Since $r(l) \uparrow \infty$, $\bar{x}(t)$ would be forced to make larger and larger jumps within an interval of $T + 1$. This leads to a contradiction since the maximum jump within any fixed time interval can be bounded using the Gronwall inequality. \square

We now state one of the main theorems of this paper.

Theorem 2. Under assumptions (A1)–(A5), almost surely, the sequence $\{x_n\}_{n \geq 0}$ generated by the SRI, given by (2), is bounded and converges to a closed, connected, internally chain transitive and invariant set of $\dot{x}(t) \in h(x(t))$.

Proof. The stability of the iterates is shown in Theorem 1. The convergence can be proved under assumptions (A1)–(A3) and the stability of the iterates in exactly the same manner as in Theorem 3.6 and Lemma 3.8 of Benaïm et al. [7]. \square

We have thus far shown that under assumptions (A1)–(A5), the SRI given by (2) is stable and converges to a closed, connected, internally chain transitive and invariant set.

3.3. Stability Theorem Under Modified Assumptions

In (A4), we assumed that $\text{Lim inf}_{c \rightarrow \infty} h_c(x)$ is nonempty for all $x \in \mathbb{R}^d$. In this section, we shall develop a stability criterion for the case when we no longer make such an assumption. In other words, we work with a modified version of assumption (A4) that we call (A6).

Modification of Assumption (A4)

Recall the following SRI:

$$x_{n+1} = x_n + a(n)[y_n + M_{n+1}] \quad \text{for } n \geq 0. \tag{8}$$

Since h_c is pointwise bounded for each $c \geq 1$, we have $\sup_{y \in h_c(x)} \|y\| \leq K(1 + \|x\|)$, where $x \in \mathbb{R}^d$ (see Proposition 1). This implies that $\{y_c\}_{c \geq 1}$, where $y_c \in h_c(x)$ has at least one convergent subsequence. It follows from the definition of upper limit of a sequence of sets (see Section 2.1) that $\text{Lim sup}_{c \rightarrow \infty} h_c(x)$ is nonempty for every $x \in \mathbb{R}^d$. It is worth noting that $\text{Lim inf}_{c \rightarrow \infty} h_c(x) \subseteq \text{Lim sup}_{c \rightarrow \infty} h_c(x)$ for every $x \in \mathbb{R}^d$. Another important point to consider is that the lower limits of sequences of sets are harder to compute than their upper limits, see Aubin and Frankowska [3] for more details.

Recall that $h_c(x) = \{y \mid cy \in h(cx)\}$, where $x \in \mathbb{R}^d$ and $c \geq 1$. Clearly, the upper limit $\text{Lim sup}_{c \rightarrow \infty} h_c(x) = \{y \mid \underline{\text{lim}}_{c \rightarrow \infty} d(y, h_c(x)) = 0\}$ is nonempty for every $x \in \mathbb{R}^d$. For $A \subseteq \mathbb{R}^d$, $\text{co}(A)$ denotes the closure of the convex hull of A , i.e., the closure of the smallest convex set containing A .

$$\text{Define } h_\infty(x) := \text{co}(\text{Lim sup}_{c \rightarrow \infty} h_c(x))$$

Below, we state the modification of assumption (A4) that we call (A6).

(A6) The DI $\dot{x}(t) \in h_\infty(x(t))$ has an attracting set $\mathcal{A} \subset B_1(0)$ and $\bar{B}_1(0)$ is a subset of some fundamental neighborhood of \mathcal{A} .

Note that in (A4), $h_\infty(x) := \overline{\text{Lim inf}_{c \rightarrow \infty} h_c(x)}$ while in (A6), $h_\infty(x) := \text{co}(\text{Lim sup}_{c \rightarrow \infty} h_c(x))$. In this section, we shall work with this new definition of h_∞ .

Proposition 2. h_∞ is a Marchaud map.

Proof. From the definition of h_∞ , it follows that $h_\infty(x)$ is convex, compact for all $x \in \mathbb{R}^d$ and h_∞ is pointwise bounded. It is left to prove that h_∞ is an upper-semicontinuous map.

Let $x_n \rightarrow x$, $y_n \rightarrow y$ and $y_n \in h_\infty(x_n)$ for all $n \geq 1$. We need to show that $y \in h_\infty(x)$. We present a proof by contradiction. Since $h_\infty(x)$ is convex and compact, $y \notin h_\infty(x)$ implies that there exists a linear functional on \mathbb{R}^d ; say, f such that $\sup_{z \in h_\infty(x)} f(z) \leq \alpha - \epsilon$ and $f(y) \geq \alpha + \epsilon$ for some $\alpha \in \mathbb{R}$ and $\epsilon > 0$. Since $y_n \rightarrow y$, there exists $N > 0$ such that for all $n \geq N$, $f(y_n) \geq \alpha + \epsilon/2$. In other words, $h_\infty(x) \cap [f \geq \alpha + \epsilon/2] \neq \emptyset$ for all $n \geq N$. We use the notation $[f \geq a]$ to denote the set $\{x \mid f(x) \geq a\}$. For the sake of convenience, let us denote the set $\text{Lim sup}_{c \rightarrow \infty} h_c(x)$ by $A(x)$, where $x \in \mathbb{R}^d$. We claim that $A(x_n) \cap [f \geq \alpha + \epsilon/2] \neq \emptyset$ for all $n \geq N$. We prove this claim later, for now, we assume that the claim is true and proceed. Pick $z_n \in A(x_n) \cap [f \geq \alpha + \epsilon/2]$ for each $n \geq N$. It can be shown that $\{z_n\}_{n \geq N}$ is norm bounded, and hence contains a convergent subsequence, $\{z_{n(k)}\}_{k \geq 1} \subseteq \{z_n\}_{n \geq N}$. Let $\lim_{k \rightarrow \infty} z_{n(k)} = z$. Since $z_{n(k)} \in \text{Lim sup}_{c \rightarrow \infty} (h_c(x_{n(k)}))$, $\exists c_{n(k)} \in \mathbb{N}$ such that $\|w_{n(k)} - z_{n(k)}\| < 1/n(k)$, where $w_{n(k)} \in h_{c_{n(k)}}(x_{n(k)})$. We choose the sequence $\{c_{n(k)}\}_{k \geq 1}$ such that $c_{n(k+1)} > c_{n(k)}$ for each $k \geq 1$.

We have the following: $c_{n(k)} \uparrow \infty$, $x_{n(k)} \rightarrow x$, $w_{n(k)} \rightarrow z$ and $w_{n(k)} \in h_{c_{n(k)}}(x_{n(k)})$ for all $k \geq 1$. It follows from assumption (A5) that $z \in h_\infty(x)$. Since $z_{n(k)} \rightarrow z$ and $f(z_{n(k)}) \geq \alpha + \epsilon/2$ for each $k \geq 1$, we have that $f(z) \geq \alpha + \epsilon/2$. This contradicts the earlier conclusion that $\sup_{z \in h_\infty(x)} f(z) \leq \alpha - \epsilon$.

It remains to prove that $A(x_n) \cap [f \geq \alpha + \epsilon/2] \neq \emptyset$ for all $n \geq N$. If this were not true, then $\exists \{m(k)\}_{k \geq 1} \subseteq \{n \geq N\}$ such that $A(x_{m(k)}) \subseteq [f < \alpha + \epsilon/2]$ for all k . It follows that $h_\infty(x_{m(k)}) = \text{co}(A(x_{m(k)})) \subseteq [f \leq \alpha + \epsilon/2]$ for each $k \geq 1$. Since $y_{n(k)} \rightarrow y$, $\exists N_1$ such that for all $n(k) \geq N_1$, $f(y_{n(k)}) \geq \alpha + 3\epsilon/4$. This is a contradiction. \square

We are now ready to state the second stability theorem for an SRI given by (8) under a modified set of assumptions. We retain assumptions (A1)–(A3), replace (A4) by (A6), and finally, in (A5), we let $h_\infty(x) := \text{co}(\text{Lim sup}_{c \rightarrow \infty} h_c(x))$. We state the theorem under these updated set of assumptions.

Theorem 3 (Stability Theorem for DI #2). Under assumptions (A1)–(A3), (A5) (where we set $h_\infty(x) := \text{co}(\text{Lim sup}_{c \rightarrow \infty} h_c(x))$) and (A6), almost surely the sequence $\{x_n\}_{n \geq 0}$ generated by the SRI, given by (8) is bounded and converges to a closed, connected, internally chain transitive invariant set of $\dot{x}(t) \in h(x(t))$.

Proof. The statements of Lemmas 1–5 hold true even when $h_\infty := \text{co}(\text{Lim sup}_{c \rightarrow \infty} h_c(x))$ and (A5) is interpreted as explained earlier. The stability of the iterates can be proven in an identical manner to the proof of Theorem 1. Next, we invoke Theorem 3.6 and Lemma 3.8 of Benaïm et al. [7] to conclude that the iterates converge to a closed, connected, internally chain transitive and invariant set of $\dot{x}(t) \in h(x(t))$. \square

Remark 1. Assumptions (A4) and (A6) required that $\dot{x}(t) \in h_\infty(x(t))$ have an attractor set inside $B_1(0)$ (the open unit ball). Further, it required $\bar{B}_1(0)$ to be in its fundamental neighborhood. Note that $h_\infty(x)$ is defined as $\text{Lim inf}_{c \rightarrow \infty} h_c(x)$ when using (A4) and it is defined as $\text{co}(\text{Lim sup}_{c \rightarrow \infty} h_c(x))$ when using (A6). Consider the following generalization of (A4)/(A6).

(A4)'/(A6)': $\dot{x}(t) \in h_\infty(x(t))$ has an attractor set \mathcal{A} such that $\mathcal{A} \subseteq B_a(0)$ and $\bar{B}_a(0)$ is a subset of its fundamental neighborhood, where $0 \leq a < \infty$.

Note that a could be greater than 1, further, since \mathcal{A} is compact by definition, a is finite. A sufficient condition for (A4)'/(A6)' is when \mathcal{A} is a globally attracting, Lyapunov stable set associated with $\dot{x}(t) \in h_\infty(x(t))$. In this case any compact set is a fundamental neighborhood of \mathcal{A} .

At the beginning of Section 3, we constructed the rescaled trajectory by projecting onto the unit ball around the origin. To use (A4)'/(A6)', we build the rescaled trajectory by projecting onto $\bar{B}_a(0)$ instead. We can modify the proofs such that the statements of Theorems 2 and 3 remain true under assumptions (A1)–(A3), (A4)'/(A6)', and (A5).

Remark 2. The advantage of using (A4)'/(A6)' is that one can conclude the stability of the iterates by merely possessing the knowledge that the associated DI of the infinity system has a global attractor set. Consider the following trivial example of a SGD algorithm with linear gradient function of the form $-(Ax + b)$. The corresponding infinity system, $\dot{x}(t) = -Ax$, is clearly “related” to the associated o.d.e. $\dot{x}(t) = -(Ax + b)$. Specifically, if there was a unique global minimizer, then both the aforementioned o.d.e.’s have a global attractor, which, in turn, implies the stability of the iterates as discussed before. This trivial example also illustrates a finer point that h_∞ and h could be related, hence information about h could help us ascertain if (A4)'/(A6)' is satisfied. Whenever possible, one could also construct Lyapunov functions to ascertain the same. While we did not consider Lyapunov-type conditions for stability, it would be interesting to extend the Lyapunov-type stability conditions developed for SREs by Andrieu et al. [1] to include SRIs.

4. Extensions to the Stability Theorem of Borkar and Meyn

We begin this section by listing the assumptions (see Section 2 of Borkar and Meyn [10]) and statement of the Borkar-Meyn theorem (See Section 2.1 of Borkar and Meyn [10]). The notations used are consistent with those of Equation (1).

(BM1) (i) The function $h: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is Lipschitz continuous, with Lipschitz constant L . There exists a function $h_\infty: \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $\lim_{c \rightarrow \infty} (h(cx)/c) = h_\infty(x)$ for each $x \in \mathbb{R}^d$.

(ii) $h_c \rightarrow h_\infty$ uniformly on compacts, as $c \rightarrow \infty$.

(iii) The o.d.e. $\dot{x}(t) = h_\infty(x(t))$ has the origin as the unique globally asymptotically stable equilibrium.

(BM2) $\{a(n)\}_{n \geq 0}$ is a scalar sequence such that $a(n) \geq 0$, $\sum_{n \geq 0} a(n) = \infty$ and $\sum_{n \geq 0} a(n)^2 < \infty$. Without loss of generality, we also let $\sup_n a(n) \leq 1$.

(BM3) $\{M_n\}_{n \geq 1}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_n := \sigma(x_0, M_1, \dots, M_n)$, $n \geq 0$. Thus $E[M_{n+1} | \mathcal{F}_n] = 0$ a.s., $\forall n \geq 0$. $\{M_n\}$ is also square integrable with $E[\|M_{n+1}\|^2 | \mathcal{F}_n] \leq L(1 + \|x_n\|^2)$, for some constant $L > 0$. Without loss of generality, assume that the same constant, L , works for (BM1)(i) and (BM3).

Theorem 4 (Borkar-Meyn Theorem). *Suppose (BM1)–(BM3) hold. Then, $\sup_n \|x_n\| < \infty$ almost surely. Further, the sequence $\{x_n\}$ converges almost surely to a (possibly sample-path dependent) compact connected internally chain transitive invariant set of $\dot{x}(t) = h(x(t))$.*

In what follows we illustrate a weakening of (BM1)–(BM3) stated above using Theorems 2 and 3. Note that (BM2) is the standard stepsize assumption, while (BM3) is the assumption on the martingale difference noise; we endeavor to weaken (BM1).

4.1. Superfluity of (BM1)(ii) As a Consequence of Theorem 2

In this section, we discuss in brief how the Borkar-Meyn theorem (Theorem 4) can be proven under (BM1)(i), (iii), (BM2), and (BM3). In other words, we show that (BM1)(ii) is superfluous. In this direction, we begin by showing the following: A recursion given by (1) satisfies (BM1)(i), (iii), (BM2), and (BM3) \Rightarrow (1) satisfies (A1)–(A5) of Section 2.2. The following implications are straightforward: (BM1)(i), (iii) \Rightarrow (A1), and (A4); (BM2) \Rightarrow (A2); (BM3) \Rightarrow (A3). We now show (BM1)(i), (iii) \Rightarrow (A5). Given $x_n \rightarrow x$, $c_n \uparrow \infty$ and $h_{c_n}(x_n) \rightarrow y$, we need to show $y = h_\infty(x)$. We have the following:

$$\|h_{c_n}(x_n) - h_\infty(x)\| \leq \|h_{c_n}(x_n) - h_{c_n}(x)\| + \|h_{c_n}(x) - h_\infty(x)\|.$$

If h is Lipschitz with constant L , then it can be shown that h_c ($h_c: x \mapsto h(cx)/c$, $x \in \mathbb{R}^d$) is Lipschitz, for every $c \geq 1$, with the same constant. Further, $h_{c_n}(x) \rightarrow h_\infty(x)$ as $c_n \uparrow \infty$. Taking limits ($c_n \uparrow \infty$) on both sides of the above equation gives $\lim_{c_n \uparrow \infty} h_{c_n}(x_n) = h_\infty(x)$ as required. Since (A1)–(A5) are satisfied, it follows from Theorem 2 that an SRE satisfying (BM1)(i), (iii), (BM2), (BM3) is stable and converges to a closed, connected, internally chain transitive and invariant set of $\dot{x}(t) = h(x(t))$ (Theorem 4).

We discuss in brief how we work around using (BM1)(ii) in proving the Borkar-Meyn theorem. The notations used herein are consistent with those found in Chapter 3 of Borkar [11]. We list a few below for easy reference.

1. $\phi_n(\cdot, x)$ denotes the solution to $\dot{x}(t) \in h_{r(n)}(x(t))$ with initial value x .
2. $\phi_\infty(\cdot, x)$ denotes the solution to $\dot{x}(t) \in h_\infty(x(t))$ with initial value x .
3. $x^n(t)$, $t \in [0, T]$ denotes the solution to $\dot{x}^n(t) = h_{r(n)}(\hat{x}(T_n + t))$ with initial value $x^n(0) = \hat{x}(T_n)$. Then, $x^n(t) = \phi_n(t, \hat{x}(T_n))$, $t \in [0, T]$.

In proving the Borkar-Meyn theorem as outlined in Borkar and Meyn [10] (BM1)(ii) is used to show that for large values of $r(n)$, $\phi_n(t, \hat{x}(T_n))$ is “close” to $\phi_\infty(t, \hat{x}(T_n))$, $t \in [0, T]$. In this paper, we deviate from Borkar and Meyn [10] in the definition of $x^n(t)$, $t \in [0, T]$, here $x^n(\cdot)$ denotes the solution up to time T of $\dot{x}^n(t) = \hat{y}(T_n + t) = h_{r(n)}(\hat{x}([T_n + t]))$ with $x^n(0) = \hat{x}(T_n)$, where $[\cdot]$ is defined in Lemma 5. In other words, we have the following:

$$x^n(t) = \hat{x}(T_n) + \sum_{l=0}^{k-1} \int_{t(m(n)+l)}^{t(m(n)+l+1)} \hat{y}(z) dz + \int_{t(m(n)+k)}^t \hat{y}(z) dz.$$

For $t \in [t_n, t_{n+1})$, $\hat{y}(t)$ is a constant and equals $\hat{y}(t_n)$. We get the following:

$$x^n(t) = \hat{x}(T_n) + \sum_{l=0}^{k-1} a(m(n) + l) h_{r(n)}(\hat{x}([t(m(n) + l)])) + (t - t(m(n) + k)) h_{r(n)}(\hat{x}([t(m(n) + k)])).$$

The proof now proceeds along the lines of Section 3.2, i.e., Lemmas 1–5 and Theorem 1; we essentially show the following: If $r(n) \uparrow \infty$, then the T -length trajectories given by $\{x^n(\cdot)\}_{n \geq 0}$ have $\phi_\infty(x, t)$, $t \in [0, T]$, as the limit point in $C([0, T], \mathbb{R}^d)$, where $x \in \bar{B}_1(0)$. This is proven in Lemmas 4 and 5, the proofs of which do not require (BM1)(ii).

4.2. Further Weakening of (BM1) as a Consequence of Theorem 3

In this section, we use the second stability theorem (Theorem 3) to answer the following question: If $\lim_{c \rightarrow \infty} h_c(x)$ does not exist for all $x \in \mathbb{R}^d$, then what are the sufficient conditions for the stability and convergence of the algorithm?

Taking our cue from assumption (A6), we replace (BM1) with the following assumption, call it (BM4).

(BM4)(i) The function $h: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is Lipschitz continuous, with Lipschitz constant L . Define the set-valued map, $h_\infty(x) := \text{co}(\text{Lim sup}_{c \rightarrow \infty} \{h_c(x)\})$, where $x \in \mathbb{R}^d$.

Note that $\text{Lim sup}_{c \rightarrow \infty} \{h_c(x)\} = \{y \mid \underline{\lim}_{c \rightarrow \infty} \|h_c(x) - y\| = 0\}$.

(BM4)(ii) $\dot{x}(t) \in h_\infty(x(t))$ has an attracting set, \mathcal{A} , with $\bar{B}_1(0)$ as a subset of its fundamental neighborhood. This attracting set is such that $\mathcal{A} \subseteq B_1(0)$.

Observe that $\text{Lim sup}_{c \rightarrow \infty} \{h_c(x)\} = \lim_{c \rightarrow \infty} h_c(x)$ when $\lim_{c \rightarrow \infty} h_c(x)$ exists. Recall the definition of Lim sup, the upper limit of a sequence of sets, from Section 2.1. It can be shown that if a recursion given by (1) satisfies assumptions (BM1)(i) and (BM1)(iii), then it also satisfies (BM4)(i), (ii). Assumption (BM4) unifies the two possible cases: when the limit of h_c , as $c \rightarrow \infty$, exists for each $x \in \mathbb{R}^d$ and when it does not.

We claim that a recursion given by (1), satisfying assumptions (BM2), (BM3), and (BM4) will also satisfy (A1)–(A3), (A6), and (A5) (see Section 3.3). From Theorem 3, it follows that the iterates are stable and converge to a closed, connected, internally chain transitive and invariant set of $\dot{x}(t) = h(x(t))$. The following generalization of the Borkar-Meyn theorem is a direct consequence of Theorem 3.

Corollary 2 (Generalized Borkar-Meyn Theorem). *Under assumptions (BM2), (BM3), and (BM4), almost surely the sequence $\{x_n\}_{n \geq 0}$ generated by the stochastic recursive Equation (1), is bounded and converges to a closed, connected, internally chain transitive and invariant set of $\dot{x}(t) = h(x(t))$.*

Proof. Assumptions (A1)–(A3), and (A6) follow directly from (BM2), (BM3) and (BM4). We show that (A5) is also satisfied. Let $c_n \uparrow \infty$, $x_n \rightarrow x$, $y_n \rightarrow y$ and $y_n \in h_{c_n}(x_n)$ (here $y_n = h_{c_n}(x_n)$), $\forall n \geq 1$. It can be shown that $\|h_{c_n}(x_n) - h_{c_n}(x)\| \leq L\|x_n - x\|$. Hence we get that $h_{c_n}(x_n) \rightarrow y$. In other words, $\underline{\lim}_{c \rightarrow \infty} \|h_c(x) - y\| = 0$. Hence we have $y \in h_\infty(x)$. The claim now follows from Theorem 3. \square

5. Applications: The Problem of Approximate Drifts and SGD

5.1. The Problem of Approximate Drifts

Let us recall the standard SRE:

$$x_{n+1} = x_n + a(n)(h(x_n) + M_{n+1}), \quad (9)$$

where $h: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is Lipschitz continuous, $\{a(n)\}_{n \geq 0}$ is the stepsize sequence and $\{M_n\}_{n \geq 1}$ is the noise sequence.

The function h is colloquially referred to as the drift. In many applications the drift function cannot be calculated accurately. This is referred to as the approximate drift problem. For more details, the reader is referred to Chapter 5.3 of Borkar [11]. Suppose the room for error is at most $\epsilon (> 0)$, then such an algorithm can be characterized by the following SRI:

$$x_{n+1} = x_n + a(n)(y_n + M_{n+1}), \quad (10)$$

where $y_n \in h(x_n) + \bar{B}_\epsilon(0)$ is an estimate of $h(x_n)$ and $\bar{B}_\epsilon(0)$ is the closed ball of radius ϵ around the origin. We define a new set-valued map called the approximate drift by $H(x) := h(x) + \bar{B}_\epsilon(0)$ for each $x \in \mathbb{R}^d$. In the following discussion, we assume that $\epsilon \geq 0$. When $\epsilon = 0$, the approximate drift algorithm described by (10) is really the SRE given by (9).

In this section, we show the following: If (9) satisfies (BM2), (BM3), and (BM4) then the corresponding approximate drift version given by (10) satisfies (A1)–(A5). For details on (BM2) and (BM3), see Section 4.1; see Section 4.2 for (BM4). We then invoke Theorem 3 to conclude that the iterates converge to a closed, connected, internally chain transitive and invariant set associated with $\dot{x}(t) \in h(x(t)) + \bar{B}_\epsilon(0) (= H(x(t)))$.

For the remainder of this section, it is assumed that (9) satisfies (BM2), (BM3), and (BM4).

Proposition 3. *$H(x) = h(x) + \bar{B}_\epsilon(0)$ is a Marchaud map. Further, recursion (10) satisfies (A1), (A2), and (A3).*

Proof. Since $\bar{B}_\epsilon(0)$ is convex and compact, it follows that $H(x)$ is convex and compact for each $x \in \mathbb{R}^d$. Fix $x \in \mathbb{R}^d$ and $y \in H(x)$, then $\|y\| \leq \|h(x)\| + \epsilon$ and $\|y\| \leq \|h(0)\| + L\|x - 0\| + \epsilon$ since h is Lipschitz continuous with Lipschitz constant L . If we set $K := (\|h(0)\| + \epsilon) \vee L$, then we get $\|y\| \leq K(1 + \|x\|)$. This shows that H is pointwise bounded. To show the upper semicontinuity of H , assume $\lim_{n \rightarrow \infty} x_n = x$, $\lim_{n \rightarrow \infty} y_n = y$, and $y_n \in H(x_n)$ for each $n \geq 1$. For all $n \geq 1$, $y_n = h(x_n) + z_n$ for some $z_n \in \bar{B}_\epsilon(0)$. Further, $h(x_n) \rightarrow h(x)$ as $x_n \rightarrow x$. Since $\{y_n\}_{n \geq 1}$ and $\{h(x_n)\}_{n \geq 1}$ are convergent sequences, $\{z_n\}_{n \geq 1}$ is also convergent. Let $z := \lim_{n \rightarrow \infty} z_n$; z is such that $z \in \bar{B}_\epsilon(0)$ since $\bar{B}_\epsilon(0)$ is compact. Taking limits on both sides of $y_n = h(x_n) + z_n$, we get $y = h(x) + z$. Thus $y \in H(x)$.

Since (10) is assumed to satisfy (BM2) and (BM3), it trivially follows that it satisfies (A2) and (A3). \square

Before showing that (10) satisfies (A4), we construct the following family of set-valued maps:

$$H_c(x) := \left\{ \frac{h(cx)}{c} + \frac{y}{c} \mid y \in \bar{B}_\epsilon(0) \right\}, \tag{11}$$

where $c \geq 1$ and $x \in \mathbb{R}^d$. In other words, $H_c(x) = h_c(x) + \bar{B}_{\epsilon/c}(0)$ for each $x \in \mathbb{R}^d$.

Proposition 4. (10) satisfies (A6).

Proof. To prove this, it is enough to show that $H_\infty(x) = h_\infty(x)$, where $H_\infty(x) := \text{Lim sup}_{c \rightarrow \infty} H_c(x)$ and $h_\infty(x) := \text{Lim sup}_{c \rightarrow \infty} h_c(x)$. Since $\dot{x}(t) \in h_\infty(x(t))$ satisfies (BM4)(ii), it trivially follows that (A6) is satisfied by (10). Note that (BM4)(ii) and (A6) essentially say the same thing.

First, we show $h_\infty(x) \subseteq H_\infty(x)$ for every $x \in \mathbb{R}^d$. Let $y \in h_\infty(x)$, $\exists c_n \uparrow \infty$ such that $h_{c_n} \rightarrow y$ as $c_n \uparrow \infty$. Since $h_{c_n}(x) \in H_{c_n}(x)$, it follows from the definition of Lim sup that $y \in H_\infty(x)$. To show $H_\infty(x) \subseteq h_\infty(x)$, we start by assuming the negation, i.e., for some $x \in \mathbb{R}^d \exists y \in H_\infty(x)$ such that $y \notin h_\infty(x)$. Let $c_n \uparrow \infty$ and $y_n \in H_{c_n}(x_n)$ such that $\lim_{c_n \uparrow \infty} y_n = y$. Since $\|y_n - h_{c_n}(x_n)\| \leq \epsilon/c_n$, we have $\lim_{c_n \uparrow \infty} h_{c_n}(x_n) = y$. We have the following:

$$\|y - h_\infty(x)\| \leq \|y - h_{c_n}(x_n)\| + \|h_{c_n}(x_n) - h_{c_n}(x)\|.$$

Taking limits on both sides, we get that $\|y - h_\infty(x)\| \rightarrow 0$, i.e., $y \in h_\infty(x)$. This is a contradiction. \square

Proposition 5. (10) satisfies (A5).

Proof. Given $c_n \uparrow \infty$, $x_n \rightarrow x$, $y_n \rightarrow y$ and $y_n \in H_{c_n}(x_n) \forall n$, we need to show that $y \in H_\infty(x)$. As in the proof of Proposition 4, we have $\lim_{c_n \uparrow \infty} h_{c_n}(x_n) = y$. Since $\|h_{c_n}(x_n) - h_{c_n}(x)\| \leq L\|x_n - x\|$, we have that $\lim_{c_n \uparrow \infty} \|h_{c_n}(x_n) - h_{c_n}(x)\| = 0$ and $\lim_{c_n \uparrow \infty} h_{c_n}(x) = y$. In other words, $y \in h_\infty(x)$. In Proposition 3, we have shown that $h_\infty \equiv H_\infty$ therefore $y \in H_\infty(x)$. \square

Corollary 3. If an SRE, given by (9), satisfies (BM2), (BM3), and (BM4)(i), (ii), then the corresponding approximate drift version given by (10) is stable almost surely. In addition, it converges to a closed, connected, invariant and internally chain transitive set of $\dot{x}(t) \in H(x(t))$, where $H(x) = h(x) + \bar{B}_\epsilon(0)$.

Proof. In Propositions 3–5, we have shown that (9) satisfies (A1)–(A3), (A5), (A6); the statement now follows directly from Theorem 3. \square

Remark 3. In the context of (9), we have that h is Lipschitz and $h_c: x \mapsto h(cx)/c$. Supposing $\lim_{c \rightarrow \infty} h_c(x)$ exists for every $x \in \mathbb{R}^d$ (see (BM1)(i) in Section 4), then $\lim_{c \rightarrow \infty} h_c(x) = \text{Lim sup}_{c \rightarrow \infty} \{h_c(x)\}$. Further, $\text{Lim sup}_{c \rightarrow \infty} \{h_c(x)\}$ is nonempty for every $x \in \mathbb{R}^d$ (since $h_c(x) \leq K(1 + \|x\|)$, $c \geq 1$) even if $\lim_{c \rightarrow \infty} h_c(x)$ does not exist for some $x \in \mathbb{R}^d$. Hence the analysis of the approximate drift problem in this section is all encompassing. The aforementioned is also the reason why in Section 4.2, we define $h_\infty(x) := \text{co}(\text{Lim sup}_{c \rightarrow \infty} \{h_c(x)\})$. It may be noted that we use $\text{Lim sup}_{c \rightarrow \infty} \{h_c(x)\}$ instead of $\text{Lim sup}_{c \rightarrow \infty} h_c(x)$ since Lim sup acts on sets and h (in this context) is a function that is not set valued. Finally, in Corollary 3, if we let $\epsilon = 0$, then we may derive Corollary 2.

5.2. Stochastic Gradient Descent

Stochastic gradient descent is a gradient descent optimization technique to find the minimum set of a (continuously) differentiable function. Suppose we want to find the minimum of $F: \mathbb{R}^d \rightarrow \mathbb{R}$ for which we can run the following SRE:

$$x_{n+1} = x_n - a(n)[\nabla F(x_n) + M_{n+1}], \tag{12}$$

where $\nabla F: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is upper semicontinuous and $\|\nabla F(x)\| \leq K(1 + \|x\|) \forall x \in \mathbb{R}^d$ (pointwise bounded). $\{a(n)\}_{n \geq 0}$ is the given stepsize sequence and $\{M_{n+1}\}_{n \geq 0}$ is the martingale difference noise sequence. If the assumptions of Benaïm et al. [7] are satisfied by (12), then the iterates converge to a closed, connected, internally chain transitive and invariant set of $\dot{x}(t) = -\nabla F(x(t))$, which is also the minimum set of F . In this section, we shall not distinguish between the asymptotic attracting set of $\dot{x}(t) = -\nabla F(x(t))$ and the minimum set of F .

As explained in Section 1, while implementing (12), one can only hope to calculate an approximate value of the gradient at each step. However, one has control over the ‘‘approximation error.’’ This is typical when gradient estimators with fixed perturbation parameters are used, it could also be a consequence of the inherent computational capability of the computer used to run the algorithm. In reality, one is running the following SRI:

$$x_{n+1} = x_n + a(n)[y_n + M_{n+1}], \tag{13}$$

where $y_n \in -\nabla F(x_n) + \bar{B}_\epsilon(0)$ and $\epsilon > 0$ is the ‘‘approximation error.’’ The following questions are natural:

1. Are the iterates stable?
2. If so, where do they converge?

Define the following set-valued map, $H: x \mapsto -\nabla F(x) + \bar{B}_\epsilon(0)$. As in (11), we define $H_c(x) := (-\nabla F(cx))/c + \bar{B}_{\epsilon/c}(0)$ and $H_\infty(x) := \text{Lim sup}_{c \rightarrow \infty} H_c(x) = \text{Lim sup}_{c \rightarrow \infty} \{(-\nabla F(cx))/c\}$. Recall the definition of Lim sup from Section 2.1.

Proposition 6. (13) satisfies (A1), i.e., H is a Marchaud map.

Proof. Given $x_n \rightarrow x$, $y_n \rightarrow y$ and $y_n \in H(x_n) \forall n$, we need to show that $y \in H(x)$. For each n , we have $y_n = -\nabla F(x_n) + z_n$, where $z_n \in \bar{B}_\epsilon(0)$. Since ∇F is pointwise bounded, it follows that $\{-\nabla F(x_n)\}$ is a bounded sequence. Let $\{n(m)\} \subseteq \mathbb{N}$ such that $\nabla F(x_{n(m)}) \rightarrow \nabla F(x)$, $y_{n(m)} \rightarrow y$. The subsequence $z_{n(m)} \rightarrow z$ for some $z \in \bar{B}_\epsilon(0)$, i.e.,

$$(-\nabla F(x_{n(m)}) + z_{n(m)}) \rightarrow (-\nabla F(x) + z) \in H(x). \quad \square$$

If in addition to (A1), Equation (13) also satisfies (A2), (A3), (A5), and (A6), then it follows from Theorem 3 that the iterates are stable and converge to a closed, connected, internally chain transitive and invariant set of $\dot{x}(t) \in (-\nabla F(x(t)) + \bar{B}_\epsilon(0))$.

Suppose F has the quadratic form $x^T A x + B x + c$, where A is a positive definite matrix, B is some matrix and c is some vector. Then, it can be shown that (A1), (A2), (A3), (A5), and (A6) are satisfied by (13) and the iterates are stable and converge to a closed, connected, internally chain transitive and invariant set of $\dot{x}(t) \in -(Ax(t) + B) + \bar{B}_\epsilon(0)$. If the comments in Remark 1 are incorporated, i.e., we use (A6)' instead of (A6), then matrix A need not be positive definite anymore.

For the purpose of this discussion, assume that ∇F is Lipschitz continuous. The graph of a set-valued map $H: \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ is given by $\text{Graph}(H) = \{(x, y) \mid x \in \mathbb{R}^d, y \in H(x)\}$. It is easy to see that $\text{Graph}(-\nabla F + \bar{B}_\epsilon(0)) \subseteq N^{2\epsilon}(\text{Graph}(-\nabla F))$. Let us also assume that \mathcal{A} is the global attractor (minimum set of F) of $\dot{x}(t) = -\nabla F(x(t))$, then every compact subset of \mathbb{R}^d is its fundamental neighborhood. It follows from the stability of the iterates that they will remain within a compact subset; say, \mathcal{U} that may be sample-path dependent. It follows from Theorem 2.1 of Benaïm et al. [8] that for all $\delta > 0$, there exists $\epsilon > 0$ such that $\mathcal{A}^\delta \subseteq N^\delta(\mathcal{A})$ is the attractor set of $\dot{x}(t) \in -\nabla F(x(t)) + \bar{B}_\epsilon(0)$. Further, the fundamental neighborhood of \mathcal{A}^δ is \mathcal{U} itself. In other words, suppose we want to ensure convergence of the iterates to a δ -neighborhood of the minimum set \mathcal{A} , then the "approximation error" should be at most ϵ (ϵ is dependent on δ).

6. Final Discussion on the Generality of Our Framework

As explained in Section 3, we run a projective scheme to show stability. In other words, time is divided into intervals of length T ; the iterates are checked at the beginning of each time interval to see if they are outside the unit ball; all the iterates corresponding to $[T_n, T_{n+1})$ are scaled by $r(n) = \|x(T_n)\| \vee 1$, i.e., the iterates are projected onto the unit ball around the origin. For $t(m(n)) = T_n \leq t(m(n) + k) < T_{n+1}$, we have the following rescaled iterate:

$$\frac{\bar{x}(t(m(n) + k))}{r(n)} = \frac{\bar{x}(t(m(n)))}{r(n)} + \sum_{j=0}^{k-1} a(m(n) + j) \left[\frac{\bar{y}(t(m(n) + j))}{r(n)} + \frac{M_{m(n)+j+1}}{r(n)} \right].$$

In the above, $\bar{y}(t(m(n) + j))/r(n) \in h_{r(n)}(\bar{x}(t(m(n) + j))/r(n))$. Since we have to worry about $r(n)$ running off to infinity, it is natural to define $h_\infty(x)$ to include all accumulation points of $\{h_c(x) \mid c \geq 1, c \rightarrow \infty\}$. This is precisely what the Lim sup function (see Section 2.1) allows us to do. In Lemma 5, it was shown that the scaled iterates track a solution to $\dot{x}(t) \in h_\infty(x(t))$ provided the original iterates are unstable, i.e., $\sup_n r(n) = \infty$. Assumptions (A4)/(A6) were never used up to this point. At this stage, it seems natural to impose restrictions on $\bar{x}(t) \in h_\infty(x(t))$ to elicit the stability of the original iterates.

As explained in Section 3.3, $\text{Lim sup}_{c \rightarrow \infty} h_c(x)$ is nonempty for every $x \in \mathbb{R}^d$ since h is pointwise bounded. Further, $h_\infty \equiv \text{co}(\text{Lim sup}_{c \rightarrow \infty} h_c)$ is shown to be Marchaud and the DI $\dot{x}(t) \in h_\infty(x(t))$ has at least one solution. Assumption (A6) is the restriction referred to in the previous paragraph that is imposed to elicit the stability of the original iterates. On a related note, if $\text{Lim inf}_{c \rightarrow \infty} h_c$ were nonempty, then we define $h_\infty \equiv \text{Lim inf}_{c \rightarrow \infty} h_c$ and check if (A4) is satisfied.

If the DI $\dot{x}(t) \in h_\infty(x(t))$ has global attractor inside $B_1(0)$, then this is a sufficient condition for (A6) to hold, it then follows from Theorem 3 that the original iterates are stable and converge to a closed, connected, internally chain transitive set associated with $\dot{x}(t) \in h(x(t))$. More generally, in lieu of Remark 1, it is sufficient that the DI has some global attractor, not necessarily inside the unit ball, since (A6)' will then hold. This, in turn, implies stability.

In case of the original Borkar-Meyn assumptions, (BM1)(i), (ii) (see Section 4) needed to be checked even before we could define h_∞ , while in our case, we do not need any extra assumptions to define h_∞ . As explained before, constructing a global Lyapunov function for h_∞ is one of many sufficient conditions that guarantee (A4)'/(A6)'.

In case of Lyapunov-type conditions for stability, additional properties of the constructed global Lyapunov function need to be verified before we get stability, see Andrieu et al. [1] for more details. However, to the best of our knowledge, there are no Lyapunov-type conditions that guarantee stability of stochastic approximation algorithms with set-valued mean fields (SRI), the class of algorithms dealt with in this paper. Hence our assumptions are general and relatively easy to verify.

7. Conclusions

An extension was presented to the Borkar and Meyn theorem to include approximation algorithms with set-valued mean fields. Two different sets of sufficient conditions were presented that guarantee the “stability and convergence” of SRIs. As a consequence of Theorems 2 and 3, the original Borkar-Meyn theorem is shown to hold under weaker requirements. Further, as a consequence of Theorem 3, we obtained a solution to the “approximate drift” problem. Prior to this paper, there was no proof of stability of SGD algorithms that use constant-error gradient estimators. Hence we could only conclude that the iterates converge to a small neighborhood, say N , of the minimum set with very high probability. In Section 5.2, we used our framework to show the stability of the aforementioned algorithm, which, in turn, allowed us to conclude an almost sure convergence to \bar{N} .

An important future direction would be to extend these results to the case when the set-valued drift is governed by a Markov process in addition to the iterate sequence. For the case of stochastic approximations, such a situation has been considered in [Borkar [11], Chapter 6], where the Markov “noise” is tackled using the “natural time scale averaging” properties of stochastic approximations. Finally, it would be interesting to develop Lyapunov-type assumptions for stability of stochastic algorithms with set-valued mean fields.

References

- [1] Andrieu C, Moulines E, Priouret P (2005) Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.* 44(1):283–312.
- [2] Aubin J, Cellina A (1984) *Differential Inclusions: Set-Valued Maps and Viability Theory* (Springer, Berlin).
- [3] Aubin J, Frankowska H (1990) *Set-Valued Analysis* (Birkhäuser, Boston).
- [4] Benaïm M (1996) A dynamical system approach to stochastic approximations. *SIAM J. Control Optim.* 34(2):437–472.
- [5] Benaïm M (1999) Dynamics of stochastic approximation algorithms. *Séminaire de probabilités XXXIII* (Springer, Berlin), 1–68.
- [6] Benaïm M, Hirsch MW (1996) Asymptotic pseudotrajectories and chain recurrent flows, with applications. *J. Dynam. Differential Equations* 8(1):141–176.
- [7] Benaïm M, Sorin S, Hofbauer J (2005) Stochastic approximations and differential inclusions. *SIAM J. Control Optim.* 44(1):328–348.
- [8] Benaïm M, Sorin S, Hofbauer J (2012) Perturbations of set-valued dynamical systems, with applications to game theory. *Dynam. Games Appl.* 2(2):195–205.
- [9] Bhatnagar S, Prasad H, Prashanth L (2012) *Stochastic Recursive Algorithms for Optimization: Simultaneous Perturbation Methods*. Lecture Notes in Control and Information Sciences, Vol. 434 (Springer, London).
- [10] Borkar VS, Meyn SP (1999) The O.D.E. method for convergence of stochastic approximation and reinforcement learning. *SIAM J. Control Optim.* 38(2):447–469.
- [11] Borkar VS (2008) *Stochastic Approximation: A Dynamical Systems Viewpoint* (Cambridge University Press, Cambridge, UK).
- [12] Kushner H, Yin G (2003) *Stochastic Approximation and Recursive Algorithms and Applications* (Springer, New York).
- [13] Ljung L (1977) Analysis of recursive stochastic algorithms. *Automatic Control, IEEE Trans.* 22(4):551–575.